

# SENTIMENT ANALYSIS ON GOJEK AND GRAB USER REVIEWS USING SVM ALGORITHM BASED ON PARTICLE SWARM OPTIMIZATION

Hermanto<sup>1</sup>, Antonius Yadi kuntoro<sup>2</sup>, Taufik Asra<sup>3</sup>, Nurajijah<sup>4(\*)</sup>, Lasman Effendi<sup>5</sup>, Ridatu Ocanitra<sup>6</sup>

<sup>1</sup>Computer Technology, <sup>3,5,6</sup>Department of Electrical Engineering  
Universitas Bina Sarana Informatika, Jakarta, Indonesia  
<https://www.bsi.ac.id>

<sup>1</sup>hermanto.hmt@bsi.ac.id, <sup>3</sup>taufik.tas@bsi.ac.id, <sup>5</sup>lasman.lef@bsi.ac.id, <sup>6</sup>ridatu.rdo@bsi.ac.id

<sup>2,4</sup>Information Systems,  
STMIK Nusa Mandiri, Jakarta, Indonesia  
<https://www.nusamandiri.ac.id>

<sup>2</sup>antonius.aio@nusamandiri.ac.id, <sup>4</sup>nurajijah.nja@nusamandiri.ac.id

(\*) Corresponding Author

**Abstract**— Users of the Gojek and Grab application can provide reviews or comments about the application on Google Play. Reviews in the form of giving opinions about their satisfaction or dissatisfaction with the services provided. So with the many opinions provided, making people selective in choosing an online motorcycle taxi service provider. The application with the best review will be chosen by the community. In previous studies regarding the classification of online ojek service review using the Naïve Bayes algorithm, C.45 and Random Forest produced an unsatisfactory accuracy of 69.18% at the highest value. This study aims to determine the extent of the analysis of Gojek and Grab application user reviews based on user comments by classifying negative and positive reviews with a higher level of accuracy than previous studies so that applications with the best reviews can be known for public consideration in using the application's services. The method used for data review classification is using the Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO). The test results on the Grab application review get the highest accuracy results in the amount of 73.09% with AUC value = 0.804, while for the test results on the application review Gojek get an accuracy value of 65.59% and AUC value = 0.680.

**Keywords:** Sentiment Analysis, Gojek and Grab Reviews, Support Vector Machine, Particle Swarm Optimization.

**Intisari**— Para pengguna aplikasi Gojek dan Grab dapat memberikan review atau komentar tentang aplikasi tersebut di Google Play. Review berupa memberikan opini mengenai kepuasan atau ketidakpuasan mereka terhadap pelayanan yang disediakan. Maka dengan banyaknya opini-opini

yang diberikan, membuat masyarakat menjadi selektif dalam memilih penyedia layanan ojek online. Aplikasi dengan review terbaik akan dipilih oleh masyarakat. Pada penelitian terdahulu mengenai klasifikasi review pelayanan ojek online menggunakan algoritma Naïve Bayes, C.45 dan Random Forest menghasilkan akurasi yang kurang memuaskan yaitu sebesar 69,18% pada nilai tertinggi. Penelitian ini bertujuan untuk mengetahui sejauh mana analisis review pengguna aplikasi Gojek dan Grab berdasarkan komentar pengguna dengan mengklasifikasi review negatif dan positif dengan tingkat akurasi lebih tinggi dari penelitian terdahulu, sehingga dapat diketahui aplikasi dengan review terbaik sebagai pertimbangan masyarakat dalam menggunakan layanan aplikasi tersebut. Metode yang digunakan untuk klasifikasi data review yaitu menggunakan Algoritma Support Vector Machine (SVM) berbasis Particle Swarm Optimization (PSO). Hasil pengujian pada review aplikasi Grab mendapatkan hasil akurasi tertinggi yaitu sebesar 73.09% dengan nilai AUC = 0,804, sedangkan untuk hasil pengujian pada review aplikasi Gojek mendapatkan nilai akurasi sebesar 65,59% dan nilai AUC = 0,680.

**Kata Kunci:** Analisis Sentimen, Review Gojek dan Grab, Support Vector Machine, Particle Swarm Optimization.

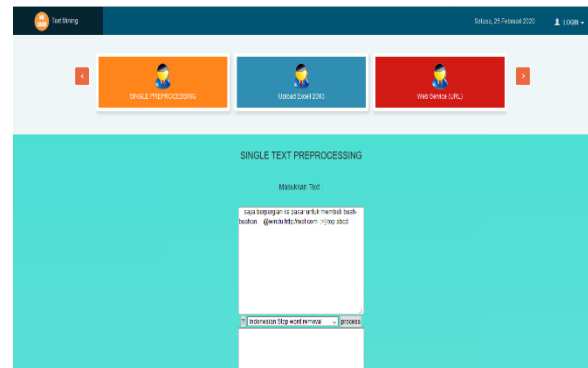
## INTRODUCTION

Ojek online provides convenience for the community in using this transportation service through a mobile application device. In Google Play, in addition to online products stores also features an assessment feature for customers to provide reviews about the advantages and

disadvantages of using the online motorcycle taxi application. Gojek and Grab user reviews may be influenced by several things that have not been a concern of both the Gojek and Grab. This might occur due to several factors that must be corrected and not yet known by Gojek and Grab. Opinions provided by the public regarding the services provided by these online motorcycle taxi services are as varied as giving opinions about the satisfaction or reduced satisfaction of the people who use these services. So with the many opinions provided, making people selective in choosing an online motorcycle taxi service provider with these conditions, from the side of the online motorcycle taxi service provider company can find out the community's satisfaction with the services provided so that it can improve continuously improving the quality of services to be provided to the customer, so research is needed to find out the application with the best review as consideration for users to use the application. Sentiment analysis is one part of text mining that is used as an analysis tool to obtain, extract and process textual data automatically to get sentiment information contained in opinion sentences (Liu et al., 2011) and can be categorized as positive or negative sentiments (Hadna et al., 2016). Research that has been done in classifying sentiments for reviews available online is Sentiment Analysis Data Social Media Facebook Comments with KNearest Neighbor (Case Study on Indonesia Express J&T Goods Expedition Services Account) (Salam et al., 2018). Previous research (Hermanto et al., 2019) on the classification of online motorcycle taxi review services using Naïve Bayes, C.45, and Random Forest algorithms obtained a less than optimal accuracy of 69.18% as the highest accuracy. In research (Nurajijah & Riana, 2019) the Support Vector Machine algorithm succeeded in defeating the accuracy of the Naïve Bayes algorithm and the Decision Tree.

Based on previous research, this study uses the Support Vector Machine (SVM) algorithm with Particle Swarm Optimization of Gojek and Grab user reviews to get higher accuracy than previous research, so that applications with the best review can be known for public consideration in using online motorcycle taxi services. The process of testing data on Rapidminer begins with the formation of a model with the data in the first part of the distribution of training data and testing data. After testing the accuracy results can be measured using a confusion matrix, and performance is measured using accuracy and AUC and will be displayed in the form of a ROC curve, of course, it can be known which online motorcycle

taxi application has the highest level of accuracy in accordance with the results of the study using the Support Vector Machine algorithm model based on Particle Swarm Optimization.



Source: <http://gataframework.com/textmining/>  
Figure 1. Display of Gata Framework Textmining tools

## MATERIALS AND METHODS

Data mining is a process of finding interesting patterns and knowledge of large amounts of data (Han et al., 2012), containing the activities of collecting, cleaning, processing, and analyzing a set of data so that with these activities a deep understanding of data can be obtained (Aggarwal, 2015).

### a. Bussines Understanding

In the stage of business understanding, carried out with an understanding of the object of research. The data that will be used in this study is a review of Gojek and Grab application users' mentors on Google Play that provide comments, to determine the status of Gojek and Grab user review comments on Google. Motivation at this stage User Gojek and Grab user review data is presented in the form text on google play that will be grouped based on the content of the discussion from each review category.

At this stage an understanding is also carried out to find the best algorithm method in order to be able to help when processing data. The algorithm used for classification in the review of Gojek and Grab application user comments is SVM with PSO. SVM is a classification method for linear and nonlinear data. PSO is a very simple optimization technique for applying and modifying several parameters (Basaria et al., 2012).

### b. Understanding data

These tapes are the process of understanding the data that will be used as

material to be studied to be carried out to the next stage, namely processing. Gojek and Grab user review data retrieve data from Google Play, where user review reviews have taken a total of 1,360 data, for Gojek user review comments amounted to 680 and Grab user comment reviews amounted to 680. From the user reviews are then grouped based on the categories of comments provided by users Gojek and Grab.

### c. Data Preparation

Preparing data to perform steps called text preprocessing, using two preprocessing applications, first using the Gata Framework which is accessed via the link <http://gataframework.com/textmining> (Figure 1) which can be used free of charge is also easy to use because do not have to create an account to use the service and continue preprocessing from Rapidminer.

Gata Framework is an alternative in pre-processing Indonesian texts combined with the Rapidminer application to process words in Indonesian, this is because in the Rapidminer application there is already a dictionary facility to change acronyms and stopwords, but it is still limited to English, Chinese, and Arabic, while for the Indonesian language is still not available. From the results of pre-processing using the Gata Framework, the data set will be pre-processed again by using the Rapidminer tools to clean the data so that the results are even better.

### d. Remove Duplicates

Remove duplicates is used to remove the same text or duplicates. This is done so that the data is not filled with the same text so that it slows the running process of software to analyze the model.

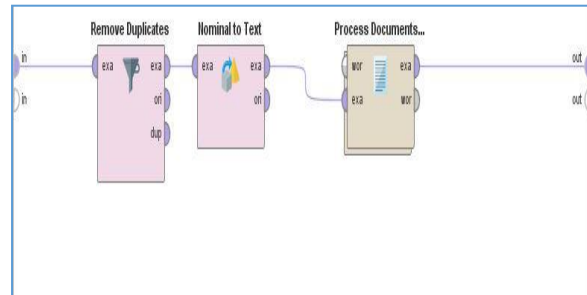
### e. Nominal to Text

Nominal to text is an operator in Rapidminer that functions to change all the numbers in the text into a text. So that the existing numbers will be considered types of text data not numeric or nominal. Figure 2 shows how the use of this operator is used in the existing process in Rapidminer.

### f. Transform Case

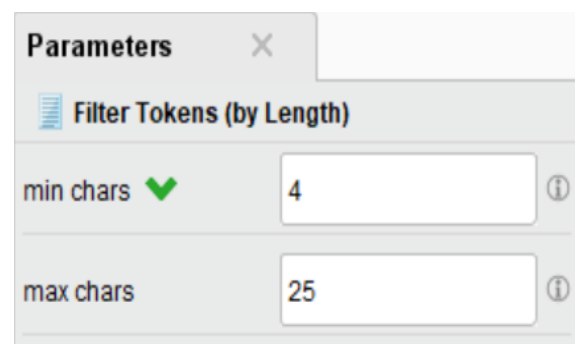
The operator used at this stage is to change the capital letters that still exist in the text will be changed to all lowercase letters. This is done so that when the process is carried out into the classification model there is uniformity of

letters and there is no error in the tokenizing process.



Source: (Hermanto et al., 2020)

Figure 2. Display Nominal to Text Process



Source: (Hermanto et al., 2020)

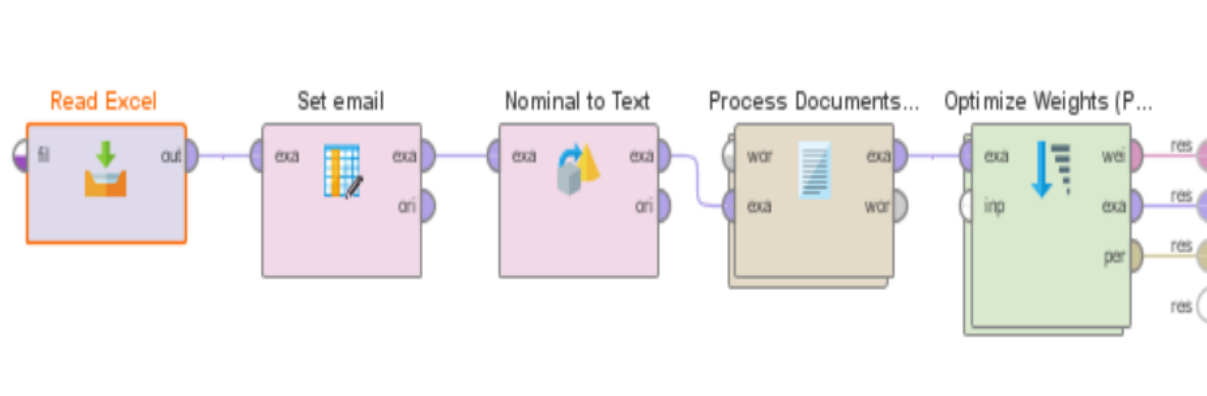
Figure 3. The Token Filter Process (By Length)

### g. Filter Token (By Length)

Filter Token is a process that exists in data preparation to eliminate a number of words (after tokenize) with a certain character length. Figure 3 explains in this study the minimum length of characters used is 4 characters and a maximum length of 25 characters. This means that words whose length is less than 4 characters and more than 25 characters will be removed. To get results like this the settings for the Parameters of this operator are done.

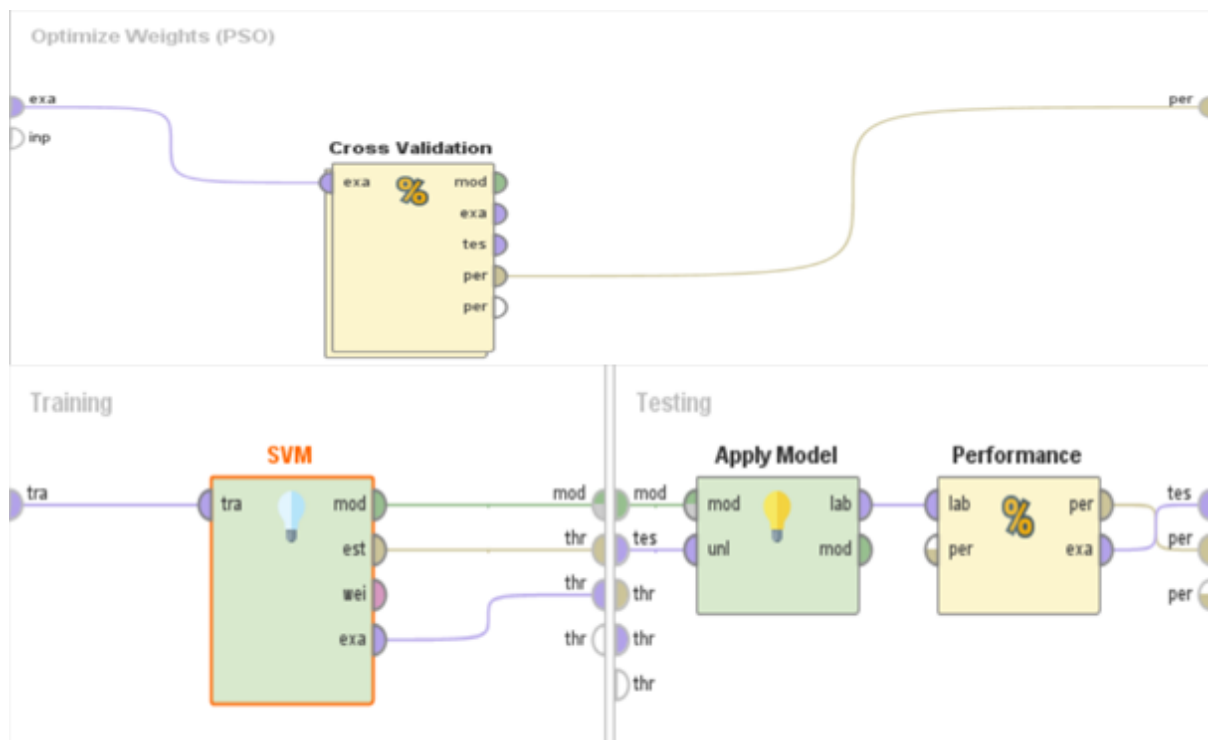
### h. Filter Stopword (Dictionary)

Next is the use of the Stopword Removal (by Dictionary) operator whose function is to remove words that are not related to the contents of the text. In the previous stage using the Gata framework text mining service has been done but there are some words that cannot be removed by the previous service because it has not been included as a word that must be deleted. Then with the Stopword Removal (by Dictionary), operator researchers can register words that should be deleted from the text.



Source: (Hermanto et al., 2020)

Figure 4. Stages of modeling



Source: (Hermanto et al., 2020)

Figure 5. Testing the Vector Machine Support Algorithm

## RESULTS AND DISCUSSION

### a. Modeling

It is the stage of selecting mining techniques by determining the algorithm to be used. The tool used is Rapidminer version 9.1. The results of the model testing conducted were analyzing Gojek and Grab user comment reviews on the Google Play Store with positive and negative comment data labels using the Vector Machine support algorithm to get the best accuracy value. The design of the Rapidminer model used can be seen in Figure 4,

beginning with entering data then setting the role set which later determines the label there and the nominal text and then processes the document. The data used is clean data that has been through preprocessing. The data is taken from the Read Excel operator, this is done because the dataset is stored in Excel (.xlsx) form. Process documents from files to convert files to documents.

### b. Validation

The settings and use of operators and parameters in the Rapidminer frameworks greatly

affect the accuracy and the model formed, more clearly testing the algorithm is explained in Figure 5. The validation process consists of training data and testing data. Then, the algorithm design model is processed into a cross-validation operator. Support vector machine, there is an algorithm calculation, then the model is applied, then entered into performance evaluation, then the accuracy and AUC results appear.

Table 1. Accuracy scores for Gojek review  
accuracy: 69,59% +/- 4,42% (micro average: 65,59%)

	true negative	true positive	class precision
pred. negative	217	122	64,01%
pred. positive	112	229	67,18%
class recall	69,96%	65,24%	

Source: (Hermanto et al., 2020)

Table 2. Accuracy scores for Grab review  
accuracy: 73,09% +/- 2,08% (micro average: 73,09%)

	true negative	true positive	class precision
pred. negative	149	39	79,26%
pred. positive	144	348	70,73%
class recall	50,85%	89,92%	

Source: (Hermanto et al., 2020)

$$\text{Acc (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} + \frac{217+229}{217+122+112+229} = \frac{446}{680} = 0,6559 \quad \dots\dots\dots(1)$$

$$\text{Acc (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} + \frac{149+348}{149+348+39+144} = \frac{497}{680} = 0,7309 \quad \dots\dots\dots(2)$$

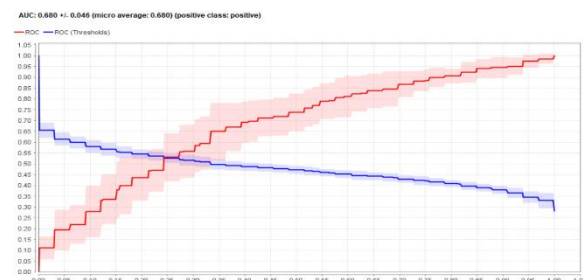
### c. Evaluation

The evaluation stage aims to determine the usability value of the model that was successfully created in the previous step. For evaluation used 10-fold cross-validation. From the results of testing the model of the two algorithms used is to produce an Accuracy (Confusion Matrix) and AUC (Area Under Curve) value. Then get the ROC graph results with the AUC (Area Under Curve) value.

Based on the results of experiments conducted using data review of Gojek and Grab application user comments on Google Play where in this experiment using the Machine Vector

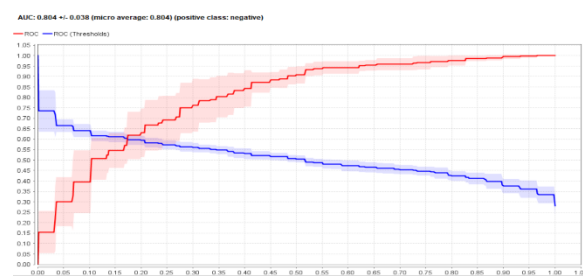
Support Algorithm using 1360 user review data, the Accuracy (Confusion Matrix) in Table 1 for Gojek and Table 2 for Grab

Based on Table 1 shows that the accuracy level of the Gojek application user review using the SVM algorithm is 65.59% which can be calculated by Formula 1. The number of True True (TP) is 229 records classified as positive and False positive (FP) is 112 records classified as negative. Next True negative (TN) 217 is classified as Negative and 122 False Negative (FN) records are classified as Negative.



Source: (Hermanto et al., 2020)

Figure 6. AUC scores on the Gojek review



Source: (Hermanto et al., 2020)

Figure 7. AUC scores on the Grab review

Table 3. Comparison of Performance Validation in Gojek and Grab Review

	Gojek	Grab
Akurasi	65,59%	73,09%
AUC	0,680	0,804

Source: (Hermanto et al., 2020)

Table 2 shows that the accuracy of the Grab application user review using the SVM algorithm is 73.09% which can be calculated by Formula 2. The number of positive True (TP) is 149 records classified as positive and False positive (FP) is 39 records classified as Negative. Next True negative (TN) 348 is classified as Negative and 144 False Negative (FN) records are classified as Negative.

From the results of model testing that has been done in Gojek review, it gets ROC graph results with Area Under Curve (AUC) value of



0.680 with accuracy performance, namely failure in Figure 6. While the results of model testing that have been done on Grab review get ROC graph results with Area values Under Curve (AUC) of 0.804 with accuracy performance that is excellent in Figure 7.

Based on the results of the analysis of each test, the results can be summarized in Table 3. The accuracy of the Grab application review testing is higher than Gojek. This study produced the highest accuracy value of 73.09%, surpassing the previous research accuracy value of only 69.18%.

### CONCLUSIONS

Evaluation and validation results are known that the accuracy value to determine that the review analysis of Gojek and Grab application user comments can be proven by the accuracy value and the AUC value of each, namely for the Gojek application with an accuracy value of 65.59% and an AUC value of 0.680, while for the Grab application, it uses an accuracy value of 73.09% and an AUC score of 0.804 with an accuracy performance that is excellent. In this study it can be seen that the level of accuracy obtained by the Grab application using the Support Vector Machine algorithm based on Particle Swarm Optimization is superior compared to the Gojek application. The application of Support Vector Machine based on Particle Swarm Optimization in this study has higher accuracy than previous research so that it can be used to provide solutions to sentiment analysis problems in the review of online application user comments.

### REFERENSI

- Aggarwal, C. C. (2015). *Data Mining The Textbook*. Springer.
- Basaria, Hasan, A. S., Hussina, B., Anantaa, I. G. P., & Zeniarj, J. (2012). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Malaysian Technical Universities Conference on Engineering & Technology 2012*, 453 – 462.
- Hadna, Shiddieqy, N. M., Santosa, P. I., & Winarno, W. W. (2016). *Literature study on comparative methods to process*.
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques. In *Data Mining*. <https://doi.org/10.1016/b978-0-12->

381479-1.00001-0

- Hermanto, H., Kuryanti, S. J., & Khasanah, S. N. (2019). Comparison of Naïve Bayes Algorithm, C4.5 and Random Forest for Classification in Determining Sentiment for Ojek Online Service. *SinKron*, 3(2), 266–274. <http://www.polgan.ac.id/jurnal/index.php/sinkron/article/view/235>
- Hermanto, Kuntoro, A. Y., Asra, T., Nurajijah, Effendi, L., & Ocanitra, R. (2020). *Final Report of Independent Research: Sentiment Analysis on Gojek And Grab User Reviews Using The Support Vector Machine Algorithm Based On Particle Swarm Optimization*.
- Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An Improved Particle Swarm Optimization for Feature Selection. *Journal of Bionic Engineering*, 8(2), 191– 200. [https://doi.org/10.1016/S1672-6529\(11\)60020-6](https://doi.org/10.1016/S1672-6529(11)60020-6)
- Nurajijah, N., & Riana, D. (2019). Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah. *Jurnal Teknologi Dan Sistem Komputer*, 7(2), 77–82. <https://doi.org/10.14710/jtsiskom.7.2.2019.77-82>
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Facebook Social Media Comment sentiment analysis with K-Nearest Neighbor (case study on Expedition service account J & T Indonesia Ekspres). *SINTAK* 2018. <https://doi.org/ISBN:978-602-8557-20-7>